A "Virtual Observatory" approach to sharing of model and simulation output, and to data-model comparisons: Outlook on CCMC support of VOs

Aaron Roberts, NASA GSFC
(Heliophysics Data and Model Consortium)
Jeffrey Hayes, NASA HQ
Darren De Zeeuw, U Michigan
(Virtual Model Observatory)

The Problem

- Many researchers want to be able to access and use data from simulations and models, often in conjunction with spacecraft data.
- Many modelers want to be able to easily share their results with others.
- The above are being done now on an *ad hoc* basis (outside of CCMC); uniformity of access and use would make everyone's research more efficient.
- CCMC is working toward such standards, but we need buy-in and a general plan.
- Additional resources are available, but we need a plan to make them effective.

NASA and More General Context

NRC Decadal Survey: DRIVE initiative

- *A relatively small, low-cost initiative, **DRIVE** provides high leverage to current and future space science research investments with a diverse set of science-enabling capabilities.*
- **D**iversify observing platforms with microsatellites and midscale ground-based assets.
- **Realize scientific potential by sufficiently funding operations and data analysis.**
- **Integrate observing platforms and strengthen ties between agency disciplines**.
- **V**enture forward with science centers and instrument and technology development.
- **E**ducate, empower, and inspire the next generation of space researchers.

Decadal Survey: Overview

- Significant progress has been made over the last decade in establishing the essential components of the solar and space physics data environment. However, to achieve key national research and applications goals*, a data environment that draws together new and archived satellite and ground-based solar and space physics data sets and computational results from the research and operations communities is needed*.

Decadal Survey: Future Goals

- Heliophysics is poised to make a natural transition from being driven predominantly by the pursuit of basic scientific understanding of physical processes towards one that must also address more operational, application-specific needs, much like terrestrial weather forecasting. *This transition requires **(1) instant unfettered access to a wide array of datasets from distributed sources in a uniform, standardized format, (2) incorporation of the results of community-developed models,** and (3) the ability to perform simulations interactively and to couple different models to track ongoing space-weather events.*
- NASA has already taken the important first step in integrating many of these datasets and tools to form the Heliophysics Data Environment (HPDE). The main objective of the HPDE is to implement a distributed, integrated, flexible data environment. HPDE modeling centers should serve as a sound foundation for a future, ***fully integrated heliophysics data and modeling center***.

VO definitions (from "VESPA")

- "The VO term covers two different meanings. It can be either "a virtual observatory": a web-based *portal* providing access to remotely distributed data resources using online forms with scientific parameters; *or "the virtual observatory": a series of* ***standards*** *and interoperable tools that can share data transparently.* In the first case, the user connects to a VO and search for data, while in the latter, the user is using tools to display data and the VO is ***the invisible machinery that allows him to work efficiently***."

VO Required Standards

- Data formats: e.g., “CDF” with specific conventions (“ISTP” “Kameleon”)
- Data product description (metadata) standards, e.g. the “SPASE” (“IMPEx” version for simulations); XML implementation
- Data transfer protocols (APIs, streaming format)
- ***All the above must be implemented (all relevant objects described) or this is useless; both agreement on standards and implementation are difficult issues***.
- (A related issue is how to deal with “big data” and server-side tools.)

How to proceed?

- A substantial part of the community should *agree on the goal*
- Standards need to be *developed/adopted*
  - CCMC has developed and implemented the Kameleon standard for output
  - CCMC is implementing SPASE/IMPEx as a metadata standard
- Community *buy-in* and implementation can lead to new *tool development* and application
- Similar standards for spacecraft data allow *data-model integration* (e.g., CCMC and "VMR" from U Michigan).

The very next step …

If you are interested in this topic, please come to the *Working Meeting* on

"Virtual Observatories and web-based information dissemination services world-wide"

Friday at 3pm

(and talk to us in the meantime)